DOCUMENT RESUME

ED 406 432 TM 026 401

AUTHOR	Smith, Mary Lee
TITLE	The Politics of Assessment: A View from the Political Culture of Arizona.
INSTITUTION	California Univ., Los Angeles. Center for the Study of Evaluation.; National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	CSE-TR-420
PUB DATE	Nov 96
CONTRACT	R117G10027
NOTE	24p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Accountability; *Culture; Educational Assessment; *Educational Change; Elementary Secondary Education; Performance Based Assessment; Policy Formation; Political Attitudes; *Political Influences; Public Policy; Qualitative Research; State Programs; *Testing Programs; Test Use
IDENTIFIERS	Arizona; *Arizona Student Assessment Program; Reform Efforts

ABSTRACT

In 1991 Arizona embarked on a program to change schools and make them more accountable for educational achievement. The instrument of reform was an assessment program known as the Arizona Student Assessment Program (ASAP), the most notable feature of which was a performance assessment that was added to an already extensive battery of state-mandated tests. This paper contains a narrative account of the events of the 4-year existence of ASAP and the research the Center for the Study of Evaluation at the University of California, Los Angeles conducted on the role of assessment in educational change. These events are explored, from 1991 until the demise of ASAP in 1995, in the context of theoretical frameworks drawn from policy studies, principally the theory of political culture. The assumption driving the use of assessment to change schools is that teachers will find ways to teach consistent with the assessment adopted. The ASAP was examined through a policy study of the beliefs of policymakers and other stakeholders, a multisite qualitative study of elementary schools in ASAP's first year of operation, and surveys and interviews of educators about ASAP impact. No sense can be made of the death of ASAP on technical or policy grounds. To understand the demise of the ASAP is to understand that such tests primarily serve political functions. The experience of the ASAP illustrates that the dynamics of wins and losses in the political arena are essential features of mandated assessment programs. (Contains 11 references.) (SLD)

CRESST

National Center for Research on Evaluation, Standards, and Student Testing

ED 406 432

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
J.C. BEER
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

The Politics of Assessment: A View From The Political Culture of Arizona

CSE Technical Report 420

Mary Lee Smith
CRESST/Arizona State University

TM026401

BEST COPY AVAILABLE

▶ UCLA Center for the Study of Evaluation

in collaboration with:

- University of Colorado
- NORC, University of Chicago
- LRDC, University of Pittsburgh
- University of California, Santa Barbara
- University of Southern California
- The RAND Corporation

The Politics of Assessment: A View From The Political Culture of Arizona

CSE Technical Report 420

Mary Lee Smith
CRESST/Arizona State University

November 1996

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

Copyright © 1996 The Regents of the University of California

The work reported herein was supported under the Educational Research and Development Center Program, cooperative agreement number R117G10027 and CFDA catalog number 84.117G, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

THE POLITICS OF ASSESSMENT:
A VIEW FROM THE POLITICAL CULTURE OF ARIZONA

Mary Lee Smith

CRESST/Arizona State University

Four years ago, the state of Arizona embarked on a new program to make schools more accountable for educational achievement and also to change them. The instrument of reform was an assessment program known as ASAP—the Arizona Student Assessment Program, the most notable feature of which was a performance assessment that was added to an already extensive battery of state-mandated tests. Four years ago, I began a program of research to investigate the policy implications underlying this program and its implications for schools. I wanted to probe the dominant hypothesis in school reform—that it is possible, and perhaps even necessary, to change the modes of assessment in order to change schools themselves. I envisioned a much different final report than the one I am now preparing. Many of the facts would have been the same, but the argument has undergone a radical change from expectations, as the Arizona assessment program has also been fundamentally altered.

This paper contains a narrative account of the events of the four-year existence of ASAP and our research on it. In addition, I attempt to make sense of these events by referring to several theoretical frameworks drawn from policy studies, principally the theory of political culture. A separate report (Smith et al., 1996) presents in greater detail the procedures and results of research as originally planned and conducted.

Key Events in the History of ASAP

Pre-1991, Arizona operated under a mandate to test in the spring of each year all common school pupils in Grades 2–12 in reading, math, and language arts, on both standardized, norm-referenced tests and the continuous uniform evaluation system (district-based, standardized, objectives-referenced tests of the Arizona Essential Skills). At that time, there was considerable opposition to standardized testing. The Center for Effective Student Evaluation had

successfully spearheaded legislation to remove first graders from the state testing mandate (Iowa Tests of Basic Skills, ITBS). The Arizona Department of Education (ADE) had contracted with Tom Haladyna and associates of Arizona State University-West to do an evaluation of ITBS and the Tests of Academic Progress. This group concluded that existing tests covered "only 26% of the Arizona Essential Skills" and confirmed the widespread discontent among educators toward the existing mandate. In addition, C. Diane Bishop, a high school math teacher, had been elected as superintendent of public instruction and head of the ADE. Her administration included such professionals as Paul Koehler and Lois Easton, who were outspoken and effective advocates for "authentic assessment," that is, assessments that fit what teachers do in classrooms, and curriculum that was more holistic and aimed toward higher order thinking and problem solving. In 1990 they mounted a campaign to convince educators to support a revision of assessment, because they believed that what gets tested is what gets taught, and teachers would revised their methods and schools their curriculum if the state renounced standardized testing in favor of performance testing. They also assumed that educators would play key roles in the planning, development, and monitoring of the testing program (their involvement would then spur professional and curriculum development by districts and teachers).

Arizona Revised Statutes 15-741 became effective in July 1991. We have pointed out elsewhere (Noble & Smith, 1994) that at least two constituencies formed a coalition to pass legislation to revise mandated testing: (a) those who, like Easton, believed that mandated standardized tests retard progress toward more holistic teaching, and (b) those who believed that schools had not been sufficiently accountable to the Arizona Essential Skills and required additional tests and procedures to correct that problem. Two such disparate senses of the problems and solutions created some incoherence at the level of the legislation that reverberated through the implementation and administration of the testing program.

When most people thought of ASAP, what they were thinking of is the Performance Test, Form D, which was only one part of the seven-part program. Form D was the only part of ASAP that incorporated the Easton ideals for assessment reform, the only part that even approached constructivism as a theory of instruction and assessment. The other parts of the ASAP program included standardized testing at three grade levels, mandated district assessment

to demonstrate district accountability to the Essential Skills, and various report cards. In addition, the legislation affirmed existing (but not previously enforced) provisions for a policy of promotion from grade to grade based on achievement of the essential skills.

ASAP as a program was then implemented by the ADE. The contract for test construction was let to Riverside, the publisher of the ITBS. Subsequently, contracts for developing scoring rubrics, and the scoring itself, were let to Riverside, Measurement Inc., and other organizations. Although the ADE conducted many workshops and made many presentations to educators about the testing program, they provided no professional development in how to teach in ways that the performance assessment suggested. Teacher training was thus left to the vagaries of the districts, some of them quite able and willing and others with little knowledge, resources, or commitments to respond.

Pilot administration of ASAP was conducted in March 1992 with results of a technical analysis reported in September 1992. The form administered was Form A, which consists of a series of items that call for students to construct responses to questions *within* the content areas of reading, math, and writing. Riverside reported acceptable levels of reliability and validity for this administration. However, they cautioned against use of ASAP pupil-level scores because reliabilities were too low for that purpose.

ASAP Form D-1 was administered in March 1993 and Form D-2 was administered in March 1994. Note that Form D differs from Form A in that the task that D entailed was *integrated* across reading, writing, and math. The scores, however, were disaggregated by content area.

ASAP as a graduation requirement came into being in January 1994 through the action of State Board of Education rule R7-2-317, which defined the level of proficiency for graduation from Grade 12. "A student shall demonstrate competency in reading, writing, mathematics, social studies and science . . . by attaining a score of 3 or 4 on each question or item of each Form A assessment [of ASAP] . . . scored with the corresponding essential skills (ASAP) generic rubric."

Technical analysis of Form D was conducted in June 1994, but the report was placed under an embargo. Riverside questioned both the reliability (for use at the pupil level) and validity (in that it failed to correlate highly with Form A) of Form D.

In November 1994, Lisa Graham won the election as Arizona Superintendent of Public Instruction, replacing Bishop, who decided not to run again. In her campaign and early days of administration, Graham advocated the introduction of marketplace reforms into public education. In other moves, she reorganized ADE, replacing staff with backgrounds in teaching and curriculum with people experienced in the private sector. In her press release on the new ADE priorities she stated plans to "refine the Essential Skills and the Arizona Student Assessment Program."

With no advance warnings and no expert or public debate, in January 1995, Graham announced that the ASAP performance test was "suspended." She explained the basis for the decision was the recent (heretofore embargoed) 1994 technical analysis that showed low correlation between Forms A and D. She was quoted by the *Arizona Republic* as saying that the suspension "won't affect the curriculum portion of ASAP, which has required teachers to change their methods of instruction. 'Instituting the program has really made a difference in the classrooms,' she said." She also was quoted as saying that teachers shouldn't worry, that ASAP would be back in 1996, and that her action was "an affirmation of ASAP and nothing less."

By May of 1995, however, "suspension" had turned into a major revision and the Arizona Student Assessment Program had transmogrified into the Arizona Student *Achievement* Program. "This is a massive change," she is quoted as saying. The new ASAP (2) now has a workplace skills component. Since "at least 50 percent of our high school students aren't college bound . . . our high schools should reflect that fact. There should be no students who aren't exposed to the workplace." In another forum, she noted that all students should know where they are headed by about junior high, and so could be directed into either a college-bound direction or a workplace direction. She also praised the work of conservative policy researcher Dennis Doyle, endorsing his educational model of holding learning constant and varying time; basing grade promotion and graduation on demonstrated mastery; rigorous, clear, measurable, standards; and the like.

To underscore the revision in standards and assessments, the ADE conducted an "Academic Summit" in October 1995. Defying the standard-setting processes used in other states (some of which required several years of development and testing), standard-setting in Arizona would be accomplished in about five days. Design teams of teachers, business leaders, and parents (but no

curriculum specialists) were commissioned to write standards in nine content areas at four levels of accomplishment. Hearings would then be conducted around the state in December, and the revised standards would be presented for State Board approval in January. If all went according to plan, requests for proposals would go out to test publishers and others to develop assessments of the approved standards in math, reading, writing, and workplace skills. The new assessments would then be developed in time for a pilot assessment in spring of 1996 and a full-blown administration in spring of 1997.

Key Events in the History of the Research Program

Having already completed a series of studies (Smith, Edelsky, Draper, Rottenberg, & Cherland, 1989) on the role of mandated testing under the pre-1991 Arizona program (universal standardized testing), I believed that the change in testing mandate called for further research. The ASAP program also offered a novel opportunity to examine the hypothesis frequently advanced by school reformers. The following paraphrase of the Resnicks goes like this: You get what you test; what you don't test, you don't get. So design assessments in the way you want students to learn, and teachers will teach that way. By altering the form of the test, one can induce teachers to accommodate their instruction to fit the test, particularly if there are consequences tied to test results. Since we know that traditional, standardized tests alter what is taught and how it is taught (curriculum narrows and teaching becomes more test-like and reductionistic), reform can be accomplished by revamping the form of assessment. By mandating performance assessment, teachers will find a way to teach in ways consistent with it, adopting the "thinking curriculum," high standards, problem solving, higher order thinking skills, and authentic, real-world, integrated problems. This is the assumption underlying the use of assessment to drive reform of schools. It is a simple assumption, perhaps behaviorist and mechanistic, but worth investigating.

To study the topic, I assembled a team of graduate students and began a series of empirical studies. The first was a policy study (Noble & Smith, 1994) that examined the beliefs and values of policy makers and other stakeholders as the legislation was passed and the ASAP program implemented at the ADE. Interviews and document analysis were the principal methods of data collection. Next, we designed and conducted a multisite, qualitative study of elementary schools operating during the initial year of ASAP implementation. We followed up

in the second year of ASAP implementation with another round of qualitative interviews with educators concerning their adaptation to the program. In addition, we conducted a survey of educators representative of educators throughout Arizona on their reactions to ASAP. This program of research was funded by the Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles. Their generous support should not be construed as extending to responsibility for the results of the study or the perspectives taken in this paper, however.

From this wealth of data, varying in approach, method, and perspective, we concluded that the measurement-driven reform hypothesis was far from convincing. Perhaps one-fifth of the schools was virtually untouched by the reform. About the same proportion had adapted wholeheartedly. In between were teachers who lacked the expertise in alternative assessment and integrated, problem-solving curriculum and pedagogy. Others disagreed with the philosophy or worked in schools driven by traditional models of teaching and testing. Still others struggled along in schools without the financial resources to devote to curriculum and professional development. Many educators were frustrated, not with the idea of performance assessment, but with this particular realization of it and many problematic features of ASAP administration and scoring. In general, our team believed at the end of this series of studies that the consequences of the ASAP mandate were uneven and perhaps distorted from program ideals, but about what one could reasonably expect of a mandate without accompanying provision for capacity development. Substantial efforts had been made by the state's educational community to respond in a professional way to the state reform.

We were never able to report that perspective, however. The data became a side-piece to the unfinished story. Political change runs faster than the policy researcher can capture it. Literally as the final pages of the report were emerging from my printer, the phone rang with news that ASAP had become history and our findings rendered moot.

Surprised as we were, we found out quickly that the movement to reform schools by reforming assessments was running into difficulty in other places as well. Ann DeVane (1995) reported at the AERA annual meeting that after four years and a $60 million investment, California abandoned CLAS. The decision was attributed to the technical weaknesses of the performance assessment, but it was really politics that killed it, according to DeVane. Analysis by Lorraine McDonnell

(1994) of the relationship of political climate and assessment policy in California, Kentucky, and North Carolina further focused my attention on the political aspects of the events leading up to the demise of ASAP. This paper now looks to political and policy theory for explanations for that demise.

Political Culture of Arizona and the Life and Death of ASAP

In their book *Culture and Educational Policy in the American States*, Marshall, Mitchell, and Wirt (1989) argued that the policy culture of a state shapes responses to national reform movements. They referred to Arizona's political culture as "traditionalistic," a culture that the economic elite (mining and agricultural interests) dominate. Noneducator interests dominate policy making over educators'. The primary policy value in the state is efficiency (tax savings) rather than excellence or equity. Education was defined as an economic function in Arizona long before it became so defined at the national level. They argued further that the professional associations in Arizona have less influence on policy making than those in other states. The data and arguments of Marshall, Mitchell, and Wirt seem to be credible in the 1990s as well. Arizona is a right-to-work state, and teachers have very little to say in a climate that systematically dismisses them.

The media also play a role in political culture. The two newspapers are owned by Dan Quayle's family. They express the values of efficiency and antiprofessionalism on a daily basis. To hear their voice alone is to believe that the teachers' "unions" are virtually dictating educational policy. They never mention an educational issue without using the term "educational establishment." With great relish, they publish the yearly results of student assessments and use these or any indicators as the source of editorial handwringing about the failure of public schools. They praise works such as those of Chubb and Moe as paragons of scientific reasoning and method. But David Berliner's (Berliner & Biddle, 1995) deconstruction of test score declines and international achievement comparisons earned him the epithet "apologist for the educational establishment." These publishers never met a choice proposal they didn't like.

The dominant view in the media matches the mood of the state government. Long before the Contract With America, the Arizona legislature was virtually nonpartisan and uni-vocal. This year, the republicans were in a majority of such dimensions that even the committee hearings on major bills barely bothered with debate—all decisions were made in caucus. In the year of the demise of ASAP, the

actions taken by Arizona government included the following. In spite of Arizona being near the bottom in spending on education, health, and social programs and near the top of the distribution of needs, the legislature failed to increase allocation even at the inflation level (in spite of this being already part of state law). Instead they passed the largest tax decrease in state history, what is known as the polluter protection bill (companies that are major sources of pollutants under investigation by environmental agencies may remain anonymous), and the "veggie hate crimes" bill (making it unlawful for anyone to defame a fruit or vegetable product). Although a full-blown school voucher program failed in the legislature, a liberal charter school legislation was passed (by the end of 1995, 50 school charters had been approved with the prospects of an additional 50 schools in 1996). In addition, the governor caused the state to sue the federal government to withdraw mandates or pay the states for implementing them. Active in the states' rights movement, he sought to avoid the federal mandates to provide school services to immigrants and to protect endangered species and fragile ecosystems.

Against this landscape of political culture, the organization of schooling struggles. Historically, districts have had more control over education than has the state government. About three-quarters of a million school children (two-thirds Caucasian) are spread across more than 200 districts of amazing variety. Some are one-school districts of a half-dozen students and others have enrollments of 25,000. Some are unified, but many are either elementary or high school districts with boundaries that cut across so many organizational lines as to make centralization improbable. Ironically, Arizona is one of the most urbanized states, if one counts the proportion of the population that lives in metropolitan centers. But the rural schools are really remote, and these include 15 districts within Indian reservations. Likewise, districts represent amazing disparities in wealth. Measured in property taxation capacity per pupil served, some districts can raise more than $50K per pupil, while others can raise nothing at all. The recent federal court case *Roosevelt District v. Bishop* declared that the differences in property taxation capacity rendered the education system inherently unequal, realizing what everyone knows by anecdotal evidence: that the roofs of some schools are literally falling around the children's heads, while other districts can afford indoor sports arenas or computers for every student. Such social and economic

disparities must be understood in relation to the potential educational effects of reform initiatives and the political uses of assessment.

The political culture framework for explaining educational policy itself builds on a "Garbage Can Model" for understanding political action. This model posits decision-making opportunities in which many problems, solutions, and policy actors are dumped in together in a kind of garbage can. These are problems in search of solutions, solutions in search of problems, and policy actors in search of both problems and solutions and their potential relationship to political prospects. The elements in the garbage can come together largely by chance, according to this model (Kingdon, 1995); that is, particular solutions get attached to particular problems largely by coincidence rather than by any inner necessity or logical coherence. This model also suggests that there are constituencies of political actors that may have alternative definitions of what constitutes a problem and what effects a policy solution is likely to have on their interests. Despite these disparate and even contradictory definitions of the situation, the groups may coalesce around a single policy solution. A policy entrepreneur may seize the opportunity to effect a coalition and attach symbols to problems and solutions that obscure the underlying contradictions in the definitions of the situation held by the various policy actors. But the resulting coalition that is based on incoherent senses of problems and solutions is unstable. The entrepreneur must act fast before the underlying incoherence surfaces. Using a garbage can model implies that the researcher collect and organize data to identify the policy entrepreneurs, policy actors, the range of definitions of the policy problems, the range of definitions of policy solutions, and the key events in policy formation and implementation, particularly as the policy gets translated through implementation hierarchies (Hall, 1994).

A garbage can model embodies an interpretivist, interactionist, and relativist view of the social world. Mucciaroni (1992) points out the limitations of the model and argues that certain structural elements, such as state political culture and policy history, act as templates to make certain policy solutions more likely than others to be attached to problems. That is, in the Arizona political culture, solutions with high values on efficiency and accountability are more likely to be pulled from the garbage can than are those of excellence and professionalism, though elements of chance and variable definitions of the situation still come into

play. The rise and demise of ASAP can be explained by attending to the elements of the garbage can and political culture models.

Pre-ASAP Policy History

The state political culture plays out in the history of accountability in Arizona schools. For many years, testing has been the dominant solution for the definition of the problem of unaccountable public institutions and public employees. Since the 1970s, Arizona school children had been the most tested in the nation. State law required that the Iowa Tests of Basic Skills or Tests of Academic Proficiency (TAP) be administered to every pupil at every grade in the spring and that the results be published by school and grade level. Teachers were almost universally dissatisfied with the state testing program. For them, the definition of the problem was the tests themselves—that standardized tests narrow the curriculum and hurt students in various ways. But this was not the dominant or successful collective definition of the situation, which held that the existing testing program actually provided too little accountability (and perhaps that teachers cheated on them), and that additional tests were needed to solve that problem. The policy history of education in Arizona (the tendency to link the assessment solution to the problem of school reform) thus influenced the direction that would be taken. At the time of the birth of ASAP, each constituent group had a different definition of the situation, however, as the following catalogue indicates.

Pre-ASAP Constituencies

Constructivist teachers. During the stage at which ASAP was introduced and discussed, many teachers advocated for it because they believed that the state-mandated, standardized tests were the principal impediment to adoption of whole language and constructivist mathematics curricula. Since ITBS and TAP employ multiple-choice forms and many districts placed importance on high scores, pedagogy tended to be reductionistic, emphasizing rote learning of basic and isolated skills. ASAP was billed as an alternative, performance-based form of assessment that would encourage integrated curriculum and pedagogy aimed at higher order problem-solving skills. ASAP was also billed as a low-stakes test that would be designed and scored with a good deal of teacher input and discretion. Thus, these teachers supported ASAP as a better alternative to the high-stakes, traditional testing program in terms of its contribution to the reform of teaching and learning.

The professional elite. This group consisted of key teacher leaders, curriculum specialists, content specialists (such as those supporting the National Council of Teachers of Mathematics math reforms), and some university faculty. Many of these individuals were high-level staff at the Arizona Department of Education or others ADE frequently called on for consultation from districts and colleges. One might also call this group the neo-liberals, for they believed in the power of government to improve and reform schools. The definition of the problem (for which ASAP would be the desired solution) held by this group was that Arizona teachers were not then focusing enough attention to the Essential Skills, the state curriculum frameworks. The Skills represented high standards, higher order thinking, and integrated problem solving and mirrored the national standards emerging from professional content specialists across the country. This group believed that the existing test mandate was part of the problem, because ITBS/TAP concentrated schools on minimal rather than high standards and failed to represent the content specializations. It was commonly said that ITBS measured only "26%" of the Essential Skills. ASAP was in turn defined as the solution to this problem because it would test in integrated format more of the Essential Skills. They also believed that ASAP would be low-stakes assessment and embody teacher input and discretion.

This group played a substantial role in informing teachers of the program, advocating its adoption, and discouraging resistance among teachers. In their advocacy, they described ASAP in its idealized image and used the term "authentic." They warned teachers that if they failed to support ASAP, even with its flaws, that the state would immediately retreat to ITBS testing.

The strong accountability group. Members of the legislature, newspaper publishers, and some ADE staff (particularly in the testing department) defined the problem of Arizona schools as lazy and incompetent teachers that needed to have their feet held to the fire by having as many tests as possible with the highest consequences attached to their results. Unlike the teachers and professional elites, they were uninterested in the form of the assessment; whether ASAP was traditional or alternative meant little or nothing. What mattered was producing more accountability at little cost, and accountability was linked to increased high-stakes testing.

The staff of the testing department was steeped in the culture of norm-referenced and criterion-referenced testing. No one had any background and

expertise in performance assessment, making the department ill-equipped to implement the Form A and Form D assessments. On the results of the performance assessments (even radically non-normal ones), they frequently tried to make interpretations more suited to norm-referenced assessments, so that the published reports of results would later be suspect. When asked to set mastery levels on the performance test scores, they simply drew the line at 75% of total score points on each assessment, ignoring its scalar properties. In addition, they tried to apply standards of reliability and validity to the results of the performance assessment, which later figured into problematic decision making at the demise of ASAP.

Testing industry. Riverside Publishers stood to lose financially when Arizona diminished its standardized testing program from all grades to three grades. It recouped some of this loss by successfully bidding on the development, administration, and scoring of the ASAP performance test. ADE obtained favorable terms in that Riverside did not bill the state for part of its development work in exchange for retaining rights to part of the product. On the other hand, development efforts were not very extensive. Form A went through a pilot and technical analysis, but we can find no record of the piloting of Form D before it was administered in spring 1993. The extremely short timeline between letting the contract for ASAP and its actual implementation precluded the kind of careful developmental work that any new technology warrants.

Antiprofessional, neo-conservative group. This vocal group believes that teaching is something less than rocket science. Speaking for this group, Governor Symington would later say that all it takes to be a good teacher is to be an educated person with an interest in teaching and a clean background check. Parents can do the job as well as professionals, and this view was operationalized in the pressure for voucher programs, charter schools, and expanded home schooling. More test scores, according to this group, can provide families with information on which to base their selection of schools and spend their vouchers. Expressed in letters to the editor, one sees the connection between big, evil government and the public schools, the linking of professional teaching with faddism, liberalism, and the capture of values from family and church. The religious right wing, which effectively opposed performance assessment in California and Kentucky, played little role in either the birth or death of ASAP.

Policy entrepreneur. Then Superintendent of Public Instruction C. Diane Bishop played a crucial role in attaching ASAP as "the solution" to the problems defined by disparate constituent groups.

Bishop came into office a democrat, an award-winning teacher advocating for teacher autonomy and higher salaries—in other words, pro-professional interests. She believed that teachers know best how students learn and what makes them fail. She increased the level of professionalism in the state Department of Education and directed department heads to serve a direct role in advising policy makers. Under her direction, the department revised the Essential Skills and attempted to incorporate the efforts of national curriculum reform groups. She seized on ASAP as her primary, perhaps her sole, policy agenda and the key to subsequent election campaigns (the superintendency is the second-highest elected office in Arizona). She spearheaded the process of legislation and beat back all attempts to weaken ASAP.[1] At the implementation stage, she managed to attach ASAP to every other policy and program and supervised the raising of stakes to be attached to its results. It has frequently been alleged that she silenced opposition and weeded out dissenters in the department. Most of her actions can be characterized as enforcing the legal mandates, centralizing authority, and standardizing practice, rather than as coalition-building and capacity development.

Minority community. We list this group as a placeholder only. Given the problematic nature of mandated assessments and their deleterious consequences for minority populations, one would expect that their advocates might play some role in the adoption and implementation of a program such as ASAP. However, this was not the case, although some members of the professional elite spoke in behalf of minority interests. In other states, minority advocacy groups have played such a role. The ADE maintained no advisory group during the implementation of the program that would monitor the relationship of testing and minority pupils.

---

[1] A teacher coalition had pressed for legislation restricting the number of hours any one pupil could be tested in his or her career. Before hearings could begin, Bishop's emissary passed a note to the chair, who then adjourned the hearing without debate, saying that the ADE had given assurance that they would take care of the problem.

Political Events

Having identified the principal actors, we now detail the major events. Two major trends are apparent: the raising of stakes on ASAP and the decline in the influence of the professionals on the assessment process.

Arizona Revised Statutes 15-741 became effective in July 1991. The legislation specifies that the State Board will (among other things):

> Adopt and implement essential skills tests that measure pupil achievement . . . of the state board adopted essential skills in reading, writing, and mathematics in grades three, eight, and twelve.
>
> Ensure that the tests are uniform across the state, scored in an objective manner, yield national comparisons, survey on "non-test indicators," require districts to submit plans for assessment of essential skills at all grade levels, publish report cards at the pupil, school, district, and state levels, and require norm-referenced, standardized tests at grades 4, 7, and 10.

In addition, the legislation affirmed existing (but not previously enforced) provisions for a policy of promotion from grade to grade based on achievement of the essential skills.

ASAP as a program was then implemented by the ADE. The contract for test construction was let to Riverside, the publisher of the ITBS. Subsequently, contracts for developing scoring rubrics, and the scoring itself, were let to Riverside, Measurement Inc., and other organizations. ADE concluded that there was not sufficient time or budget to allow teachers to contribute to the development and scoring of ASAP, thus these processes were contracted out. Some teachers served on an advisory panel and were hired to serve as scorers. In terms of staff development, ADE conducted many in-service programs on the nature of ASAP, how it would be administered and scored. But there was never any provision for professional development of teachers in how to teach in ways consistent with the performance assessment (i.e., integrated, thematic, problem-solving curriculum)—that was left to the resources and prerogatives of districts and schools.

Pilot administration of ASAP was conducted in March 1992 with results of a technical analysis reported in September 1992. The form administered was Form A, which consists of a series of items that call for students to construct responses

to questions *within* the content areas of reading, math, and writing. Riverside reported acceptable levels of reliability and validity for this administration. However, they cautioned against use of ASAP pupil-level scores because reliability was too low for that purpose. According to both testing company and ADE testing department staff, the technical adequacy of ASAP had to be demonstrated because of the need for comparability and reliability. In other words, ASAP was to be used for accountability just as the ITBS had been.

ASAP Form D-1 was administered in March 1993 (D-2 was administered in March 1994). Note that Form D differs from Form A in that the task that D entailed was *integrated* across reading, writing, and math. The scores, however, were disaggregated by content area. When D-1 scores were reported in June 1993, they were published in the same way that ITBS scores had been previously published. Bishop expressed her strong disappointment with the low scores, saying that teachers were not performing as needed or adapting properly to the new assessments. Teachers expressed their surprise and dismay at the unexpected way that the state was using ASAP.

ASAP as a graduation requirement came into being in January 1994 through the action of State Board of Education rule R7-2-317, which defined the level of proficiency for graduation from Grade 12. "A student shall demonstrate competency in reading, writing, mathematics, social studies and science . . . by attaining a score of 3 or 4 on each question or item of each Form A assessment [of ASAP] . . . scored with the corresponding essential skills (ASAP) generic rubric." This event fully institutionalized ASAP from a program that ADE promulgated to a formal state policy. The ADE also announced that it would begin to enforce the state legislation that tied grade promotion decisions to mastery of the Essential Skills, as measured by ASAP performance at Grades 3 and 8. In addition, ADE supported legislation to base district takeover decisions on the results of ASAP. The ADE staff interpreted the Goal-Setting provision of ASAP as how a district planned to increase ASAP scores, rather than as how it could provide better quality educational programs.

Technical analysis of Form D was conducted in June 1994, but the report was placed under an embargo. Riverside questioned both the reliability and validity of Form D. In their analysis, they made no attempt to correlate Forms D-1 and D-2.

State superintendency turns over in November 1994. Lisa Graham places high priority on introducing marketplace reforms into public education. In other moves, she reorganizes ADE, replacing staff with backgrounds in teaching and curriculum with business people. Even the testing coordinator is demoted. In her press release on the new ADE priorities Graham lists "refine the Essential Skills and the Arizona Student Assessment Program."

ASAP is suspended. The *Arizona Republic* announced the decision by Lisa Graham on January 21, 1995. She accounts for the decision based on the June 1994 technical analysis that showed low correlation between Forms A and D. (Since the two forms were essentially measuring two different kinds of tasks, one would not expect high correlation, however.) She was quoted as saying that the suspension "won't affect the curriculum portion of ASAP, which has required teachers to change their methods of instruction. 'Instituting the program has really made a difference in the classrooms,' she said." She also was quoted as saying that teachers shouldn't worry, that ASAP would be back in 1996, and that her action was "an affirmation of ASAP and nothing less."

Suspension becomes major revision. On May 26, 1995, the *Arizona Republic* reported that ASAP had transmogrified into the Arizona Student *Achievement* Program, and that it will subsequently be administered to fourth, eighth, and tenth graders. "This is a massive change," she is quoted as saying. The new ASAP (2) now has a workplace skills component. Since "at least 50 percent of our high school students aren't college bound . . . our high schools should reflect that fact. There should be no students who aren't exposed to the workplace."

Constituencies at the Demise of ASAP (1)

At the birth of ASAP, a number of policy actors, clumped into constituent groups, acted together to attach a particular solution to one of several perceived problems (or at least they failed to resist this attachment) in the garbage can in 1991. By 1995, what had happened to these groups that might help to explain the demise of ASAP (1)?

First, the constructivist teachers had largely abandoned ASAP as the solution to problems as they defined them. Many came to realize that ASAP was not even very constructivist. Form D did present students with interesting, real-world problems and integrated subject matter, but it was not authentic, teacher-directed, instruction-embedded assessment. These qualities had been sacrificed to

the values of objectivity and standardization. Furthermore, they realized that ASAP had added to the testing burden and the high-stakes accountability load. Strong advocates of ASAP just one year earlier, the Hilldale School whole language teachers reported in 1995, "ASAP just gets in our way." Although many teachers in our survey had positive attitudes about the ASAP performance assessment, they liked it more in the idea than in the realization (Smith et al., 1996). Many teachers also realized by that time that they had been cut out of the development and implementation process and had become the objects of change rather than the agents of change. Thus, this constituent group was no longer an effective advocate when ASAP (1) was suspended.

The professional elite had been the foremost symbolizers of ambitious, integrated, "thinking" curriculum and alternative assessments, but they never succeeded in building consensus or understanding in the public, the media, or in many schools that continued to embrace traditional education. Gradually, members of this group drifted away from ADE to district leadership positions and national reform groups. Over time, those who had been employed as consultants to the department were called less and less often. One can argue that members of this group had become aware of the discrepancies between their idealized image of what ASAP could have been, and the emerging realities. Their objections were effectively silenced in the ADE. As this group diminished in status and number in ADE, they were replaced by members of the strong accountability group or partisans of the superintendent. They never understood the constructivist nature of ASAP well enough to keep those ideals alive.

The strong accountability group: Though they never met a test they didn't like, the ASAP performance test came close. It lacked the degree of objectivity and standardization possessed by traditional, standardized tests, and it cost too much. Since constructivism was not an ideal this group was concerned about, abandoning the format was not a risk.

The testing industry: Like the strong accountability group, the testing publishers have little vested interest in the form of the assessment—only in the existence of some kind of assessment from which they can potentially profit. Bidding will soon open for development, administration, and scoring of ASAP (2), and representatives of several testing publishers and other developers have been highly visible at the Academic Summits of 1995.

The antiprofessional neo-conservatives remain as they were, not so concerned about the form of the test as they are about its uses. Antigovernment, antipublic institution sentiment has, if anything, grown stronger with the passage of permissive charter school legislation. It finds expression in the proposals to abolish school districts, abolish requirements that teachers be certified, abolish teacher training programs, and even abolish ADE.[2]

Perhaps more potent than any of the changes in constituent groups is the change in policy entrepreneur. For personal and financial reasons, Bishop decided not to seek reelection, although she was appointed as special assistant to the governor on education affairs after she switched parties and became an advocate for school vouchers. The new entrepreneur is Lisa Graham, who would fit in our category system with the antiprofessional neo-conservatives. A bright, attractive woman, she was formerly chair of the House Education Committee. Her sense of the "problem" is that schools are bureaucratic, underachieving, and not sufficiently accountable. Her sense of the "solution" consists of the introduction of free market forces—choice, supported by information that parents can use to exercise that choice. The charter school program, which is the centerpiece of her administration, is one of the most permissive in the nation relative to its requirements about who can teach (anyone who passes a background check, regardless of education, training, or certification) and what curriculum can be offered (no restrictions). In her Academic Summit, she asked that revisions of the state curriculum frameworks be "precise, measurable, easily understood by parents and the public." She is little interested in constructivist ideals of instruction and measurement and claims that ASAP (2) will be at least partly traditional in form and conducive to the qualities of comparability and objectivity. Because time and resources are short, development work will be contracted out, rather than assigned to teachers. In other antiprofessional actions, Graham dismantled ADE so that all curriculum specialists were moved to the bottom of the organizational chart or out of the department altogether. In their places are now bright, young people with backgrounds in business rather than education.

[2] The proposal suggested that ADE be abolished except for an office that would dictate assessment and reporting of data. Each school would act as a semi-autonomous agent, but budgeting, accounting, and data collection and reporting would be controlled on-line by the state department.

The Role of Politics in Assessment

Professionals and scholars of assessment tend to define state assessment programs on their own terms, that is, as instruments of reform or instruments that measure pupil achievement more or less well. We discuss among ourselves how performance assessment differs from traditional, standardized assessment on grounds of validity, reliability, consequences to the system, and the like. Yet the history of ASAP shows that the professional viewpoint is at best a partial one. No sense can be made of the demise of ASAP (1) on either technical or policy grounds. One could not rationally expect a performance assessment, a new and relatively untried technology, to achieve the same standards as the fifty-year-old experiment with multiple-choice tests. The evidence on ASAP was certainly mixed, but not so negative and utterly preliminary to justify the decision made. Nor did the state have any evidence of the failure of ASAP to achieve its reform goals. To our chagrin as policy researchers, our findings had not yet been made available by the time the decision was made. Nor did ADE attempt to evaluate for itself the impact the program had had on schools. In any case, the program had not been in place long enough for reasonable effects to be realized. Having spent considerable effort in responding to ASAP (1), many Arizona educators are now waiting anxiously to see what the new program will demand of them.

To understand the demise of ASAP is to understand that tests such as ASAP serve primarily political functions. As a political instrument, ASAP sat uneasily in the political culture that emphasizes efficiency, decentralization, accountability, and antiprofessionalism. The linking of the tests to the accountability value was achieved symbolically by the policy entrepreneur, and she was able to tie the health of the program in with her own political interests. The symbolic attachment gave way under the new policy entrepreneur, whose efforts were met with little resistance from the old constituencies favoring ASAP. The new definition of accountability was to be less associated with testing and more associated with free market mechanisms, under the new administration.

This paper has argued that mandated assessment programs are more than marks on optical scanning sheets, assignment of rubric scores to essays, or the accommodation of teachers to measurement-driven reforms. One must examine instead the dynamics of wins and losses in the political arena.

References

Berliner, D., & Biddle, B. (1995). *The manufactured crisis*. Reading, MA: Addison-Wesley.

DeVane, A. (1995, April). *Derailing CLAS: Assessment reform in California and the rhetoric of resistors*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Hall, P. (1994). *Policy as the transformation of intentions*. Unpublished manuscript, University of Missouri, Department of Sociology.

Kingdon, J. W. (1995). *Agendas, alternatives, and public policies* (2nd ed.). New York: Harper-Collins.

Marshall, C., Mitchell, D., & Wirt, F. (1989). *Culture and educational policy in the American states*. New York: Falmer.

McDonnell, L. M. (1994). *Policymakers' views of student assessment* (CSE Tech. Rep. No. 378). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Mucciaroni, G. (1992). The garbage can model and the study of policy making. A critique. *Polity, 24*, 459-482.

Noble, A. J., & Smith, M. L. (1994). Old and new beliefs about measurement-driven reform. *Educational Policy, 8*, 111-136.

Smith, M. L. (1996). *Assessment reform and reform by assessment: Research on the Arizona Student Assessment Program*. Phoenix, AZ: Southwest Educational Policy Studies.

Smith, M. L., Edelsky, C. E., Draper, K., Rottenberg, C., & Cherland, M. (1989). *The role of external testing in elementary schools*. Los Angeles: University of California, Center for the Study of Evaluation.

Smith, M. L., Noble, A. J., Heinecke, W., Seck, M., Parish, C., Cabay, M., Junker, S. C., Haag, S., Tayler, K., Safran, Y., Penley, Y., & Bradshaw, A. (1996). *Reforming schools by reforming assessment: Consequences of the Arizona Student Assessment Program* (Deliverable to OERI). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.